

# 画像認識による家庭廃棄物推定における背景除去効果の基礎検討

黄 思韵<sup>1</sup> 筒口 拳<sup>1</sup>

**概要:** 本研究では、家庭廃棄物を撮影した画像からリサイクル可能な資源物を自動識別することを目的として、撮影画像と、そこから背景を除去した画像との機械学習による推定精度比較を行った。データセットとして TrashNet を用い、VGG16 の転移学習を用いた CNN により学習を行なった上で実際に撮影した家庭廃棄物画像の分類を行った。その結果、廃棄物種別により推定精度に差異が生じているが、背景を除去する前処理を行うことで、推定精度に向上がみられることがわかった。

**キーワード:** 機械学習, CNN, 画像認識, 背景除去

## Background Removal Effect in Estimation of Household Waste by Image Recognition

SIYUN HUANG<sup>†1</sup> KEN TSUTSUGUCHI<sup>†1</sup>

**Abstract:** This study aims to automatically identify recyclable resource materials from images taken of household waste. We compared the accuracy of estimation by machine learning between the captured image and the image from which the background was removed. We used TrashNet as the dataset, trained with a CNN using VGG16 transition learning, and then classified actual images of household waste. As a result, there are differences in the accuracy of estimation by waste type, but we found that pre-processing to remove the background improved the accuracy of the estimation.

**Keywords:** machine learning, CNN, image recognition, background removal

### 1. はじめに

地球環境保全のため、そして子どもたちの未来のため、ゴミを減らし、資源を再利用することが重要である。近年は学校でタブレットが用いられることもあり、子どものときから地球環境への意識を高めるため、我々は、子どもたちでも容易に扱えるインタフェースを備え、撮影されたゴミの画像からリサイクル可能な物体を識別してその再利用先を提示する児童向けアプリケーションの開発をめざしている。その完成イメージを図 1 に示す。

本報告では、このアプリケーションの機能のうち、家庭廃棄物を撮影した画像中の被写体を識別し分類する手法において、画像の背景を除去した効果について述べる。機械学習を用いて撮影画像から廃棄物種別を分類する先行研究もあるが、本研究では画像に前処理を施すことによる精度向上を目標としている。

廃棄物の種別は自治体等によって異なることもあるため、分類カテゴリーをどのように設定するかという課題もあるが、本研究においては、既存のデータセットである TrashNet[1]を用いて CNN と転移学習を用いて学習モデルを構築し、そのカテゴリーに従って分類を試みた。また、

問題設定をこの既存データセットに対するテスト画像の背景除去の効果に絞り込み、撮影したオリジナル画像と背景を除去した画像とで分類精度の比較を行った。

背景除去には GrabCut 法[2]を用いた。その結果、対象とするカテゴリーにより差異はあるものの、背景除去による精度向上の効果が見られた。

以下、第 2 章で関連手法について説明し、第 3 章で本研究の提案手法について述べ、第 4 章で実装および実装、その結果について記述し、その後、考察や今後の課題について述べる。



図 1 アプリケーションの利用イメージ

<sup>1</sup> 崇城大学 情報学部  
Faculty of Computer & Information Sciences, Sojo University

## 2. 関連手法

画像認識の分野で広く用いられている手法に CNN (Convolutional Neural Network) がある。CNN はいくつもの深い層を持ったニューラルネットワークであり、畳み込み層 (Convolution Layer) とプーリング層 (Pooling Layer) から構成される。また、学習済みモデルを用いた転移学習による効率化も知られている[3]。

そのような背景のもと、小助川らはゴミ分別を目的として、CNN を用いて、入力画像が低画質であるモデルと高画質であるモデルを作成し、その2つのモデルの出力を足し合わせることにより、認識精度を向上させる手法を提案した[4]。小助川らの手法では単純な背景に比較して、複雑な背景を持つ画像では認識精度が低下する課題が挙げられている。

また、上野らは果物画像からの糖度推定を目的として、MLP (Multi Layer Perceptron) と CNN を結合したハイブリッドモデルを用いる手法を提案している[5]。上野らの手法では撮影画像をトリミングする手法を用いているものの、背景除去などの前処理が課題として挙げられている。

一方、画像中の物体を抽出する技術として、GraphCut やそれを発展させた GrabCut[2]、深層学習を用いた領域分割・抽出技術などが多く提案されている。

これらの状況を踏まえ、本研究においては、GrabCut による撮影画像の背景除去の効果を確認するものである。

## 3. 提案手法

本研究では、入力画像に前処理を施し、対象とする物体を背景から分離することで精度向上をめざしている。このため、以下の手順でその有効性を確認する：

1. TrashNet データセットから学習モデルを作成する。
2. オリジナル画像での精度を確認する。
3. 画像に前処理を施した画像での精度を確認する。

図2は実験の概要を示す：

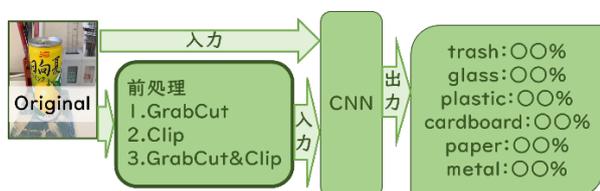


図2 実験の概要 (入力は画像)

本研究では第2章で述べた通り背景除去の効果を確認することに主眼をおくため、転移学習や CNN の構造についての詳細な設計は行わず、同一の条件で比較できるようにするために、既存のデータセット (TrashNet) を使用し学習モデルとして VGG16 を用いた。

撮影画像に対する前処理については、背景除去に加え、被写体周辺をトリミングする (本稿では clip と称する) 前処理も加え、以下の3手法で前処理を行った画像と、オリジ

ナル画像との認識結果を比較することとした：

1. GrabCut. GrabCut 法による前景領域分割処理を行う。

光の加減などの撮影状況によって対象物体がはっきり分割できないこともある。図3は段ボールの画像を GrabCut で処理した例である。

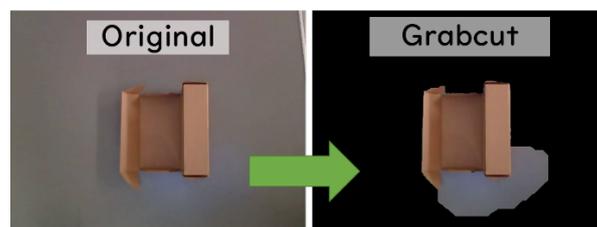


図3 GrabCut 法による領域分割処理例 (右)

2. Original&Clip. 原画像から対象物周辺の矩形を指定することによりを切り出す。矩形指定は手作業で行う。図4は原画像から対象物体を切り出した画像例である。



図4 対象物体を切り出した画像例

3. GrabCut&Clip. 1で述べた背景除去画像に対し、2で述べた切り出しを行う。図5に例を示す。

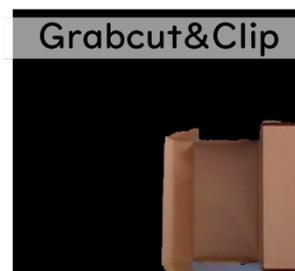


図5 GrabCut&Clip の画像例

## 4. 実装・実験

本章では学習モデルおよび実験について説明する。

### 4.1 データセット

本研究では、TrashNet[1]を用いて学習モデルを作成した。画像サイズが 531×384 ピクセルで、合計 2527 枚の画像があり、RGB3 チャンネルのカラー画像である。クラス数は6で、「glass (ビンなどのガラス)」、「cardboard (段ボール)」、「paper (紙類)」、「plastic (プラスチック)」、「metal (金属)」、「trash (その他の廃棄物)」に分類され、各画像は属するクラス名のディレクトリに配置されている。

廃棄物の公開データセットとしては TACO[6]もある。

TACO は南国のビーチからロンドンの街角といったさまざまな環境で撮影されたゴミの画像が収録されており、アノテーションも付与されているが、本研究は家庭からの廃棄物を主たる対象としているため、そのような分類に近い TrashNet の方がより適していると考え、採用した。

なお、TrashNet における「plastic」にはペットボトルが多く含まれており、容器包装プラスチックは「plastic」に分類されている画像もあれば、「trash」に分類されている画像もある。

#### 4.2 学習モデル

画像認識処理は CNN により VGG16[7] を転移学習させ、オプティマイザとして Adam と SGD の 2 つを用いて、2 種類の学習モデルを構築した。データは TrashNet の画像のうち、学習データに 80%、テスト用データに 20% を用いた。

実装および実験は Google Colaboratory 上で Keras を用いて行った。エポックが 20 で、ミニバッチが 16 とした。

Adam の場合のテストデータ正解率は 0.6772 で、SGD の場合のテストデータ正解率は 0.8399 となり、テストデータにおいては Adam より SGD の方が正しく識別できると考えられる。それぞれの学習曲線を図 6、7 に示す。学習過程において、過学習あるいは学習不足という問題は特に生じなかった。

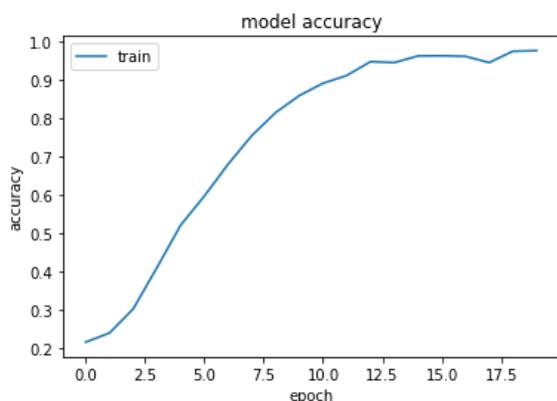


図 6 Adam の学習曲線

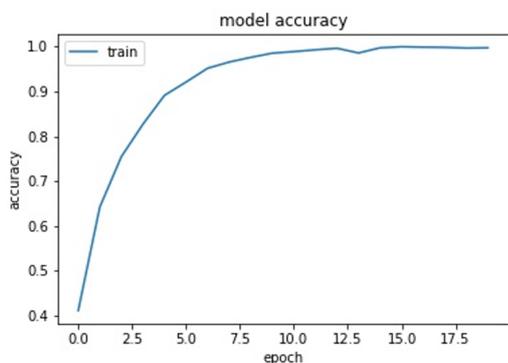


図 7 SGD の学習曲線

#### 4.3 実験

作成した 2 つのモデルに対し、独自に撮影した 60 枚の検証用画像を用いて、識別精度を確認する実験を行った。TrashNet の 6 種類のクラスに対応した、6 種類の廃棄物の画像をそれぞれ 10 枚ずつ撮影し、3 章で述べた 3 種類の前処理を加えた。したがって、オリジナル画像 60 枚、前処理画像 60 枚×3 種類の 180 枚、合計 240 枚を 4.2 節で構築した 2 つの学習モデルに入力し、識別結果の正解率を測定した。その結果を表 1 (Adam)、表 2 (SGD) に示す。

本実験においては、例えば、plastic と trash の区別などが曖昧なこともあり、上位 2 位までに含まれた場合は正解としている。

表 1 Adam の実験結果

Adam				
2位まで	org	clip-org	grabcut	clip-grabcut
plastic	0.70	0.60	0.50	0.50
cardboard	0.50	0.60	0.50	0.60
paper	0.80	0.60	0.90	0.80
trash	0.10	0.10	0.50	0.20
metal	0.00	0.20	0.10	0.50
glass	0.50	0.50	0.20	0.30

表 2 SGD の実験結果

SGD				
2位まで	org	clip-org	grabcut	clip-grabcut
plastic	0.60	0.40	0.40	0.20
cardboard	0.70	0.70	0.70	0.90
paper	0.70	0.60	0.80	0.70
trash	0.60	0.60	0.90	0.60
metal	0.20	0.20	0.10	0.30
glass	0.40	0.40	0.20	0.30

実験結果から以下のことが言える：

1. オプティマイザにかかわらず、4 つの種類 (cardboard, paper, metal, trash) において、背景除去による分類精度の向上が見られる。
2. 逆に plastic (ペットボトル), glass といった比較的透明な被写体については精度が減少している。
3. Adam においてはオリジナル画像からのクリップ処理のみの効果が見られるものがあるが (cardboard, metal), SGD においてはクリップ処理のみの効果は特にみられない。
4. 前処理の効果としては主として背景除去により、Adam は最大 40 ポイント (metal), SGD は最大 30 ポイント (trash) の精度向上がみられたことがわかった。

よって、分類種類によるものの、背景除去の効果があったと考えられる。

ここで、正解となった例と不正解となった例を図 8, 9 に示す。いずれも、画像 1, 2, 4 は GrabCut 法による前処理を施した画像で、画像 3 は GrabCut&Clip の画像例である。

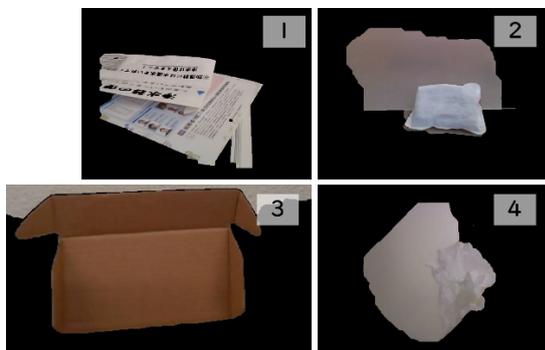


図 8 正解例 (1:paper, 2:trash, 3:cardboard, 4:trash)

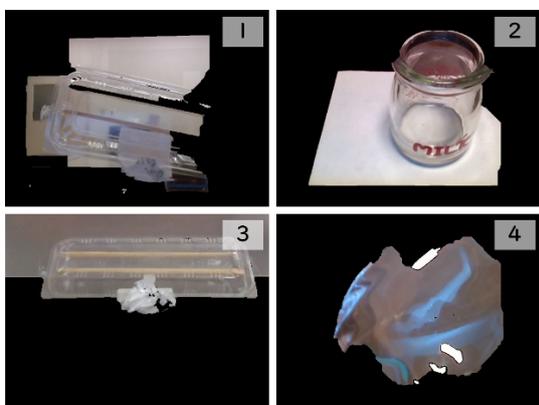


図 9 不正解例 (1,3:plastic, 2,4:glass)

## 5. 考察

本研究ではオプティマイザが Adam であり SGD であり、GrabCut 法による背景を除去した画像の認識精度が比較的向上した。しかしながら、特に精度に変化がみられないものや、低下したものもある。原因のひとつは TrashNet データセットの中で分類に曖昧なものがあった点や、検証用画像の「正解」は筆者らの主観的な判断に基づいているため、分類自体が筆者らの常識と異なる部分もあったためと考えられる。特に紙ごみなどは再利用可能な「paper」なのか、廃棄する「trash」なのかの判断は人間でも難しいことがある。また、今回、GrabCut は撮影画像に対して実施したが、実利用を考えると、撮影と同時に背景を除去できる仕組みがあると利用者の負荷を減らすことができる。

これらを踏まえると以下のような改善点が考えられる：

- TrashNet 以外のデータセット (TACO など) との混合モデルや独自画像によるモデルの構築と分類。
- 簡易な操作 (撮影の際の被写体のタップなど) による対象物体の抽出処理。

## 6. まとめと今後の課題

本研究では、地球環境保護・教育目的として、物体識別を用いた家庭廃棄物の分類システムを提案し、画像中の背景を除去することにより分類精度が向上することを確認した。TrashNet データセットを用いて CNN による学習モデルを構築し、独自の撮影画像と、それに 3 種類の異なる前処理を行った画像との分類を比較し、種別にはよるものの背景除去の効果が認められた。

しかしながら、分類クラスの妥当性や、画像に複数の対象物がある場合の選定などの課題があるため、今後は以下を進めていく予定である：

### 1. 独自データセットの構築。

混合モデルや独自画像によるデータセットを構築し、利用目的に沿った分類・ラベリングを行う。

### 2. 容易な背景除去手法の検討。

ワンタッチで対象物を指定でき、それ以外の背景を除去できるインターフェースを実装する。

### 3. 学習モデルの検討

本報告では VGG16 の転移学習を用いたが、背景除去の効果がある程度確認できたため、より有効なモデルの有無を検討していく。

## 参考文献

- [1] Gary Thung: "trashnet", <https://github.com/garythung/trashnet> (2022.12 access).
- [2] C. Rother, V. Kolmogorov and A. Blake: "'GrabCut': interactive foreground extraction using iterated graph cuts", Proceedings of SIGGRAPH 2004, pp.309-314 (2004/08).
- [3] 藤吉弘亙, 平川翼, 山下隆義: "深層学習が変えた画像認識へのアプローチ", 電子情報通信学会誌, Vol.105, No.5, pp.364-370 (2022年5月).
- [4] 小助川克也, 小泉康一, 大槻正伸: "物体認識システムを用いたゲームの開発", 情報処理学会研究報告, Vol.2019-GI-41, No.3 (2019).
- [5] 上野延典, 本田耕大, 森本正志: "ハイブリッド深層学習モデルを用いた果物画像からの糖度推定アプリケーションに関する検討", 情報処理学会研究報告, Vol.2022-CVIM-231 No.14 (2022年11月).
- [6] TACO dataset : <https://tacodataset.org/> (2022.12 access).
- [7] Karen Simonyan, Andrew Zisserman: "Very Deep Convolutional Networks for Large-Scale Image Recognition", ICLR2015 (2015).